



Display of NCBI Sequence Records

Prominent functional links and enhanced display to facilitate usage and discoveries

<http://www.ncbi.nlm.nih.gov/nuccore/> & <http://www.ncbi.nlm.nih.gov/protein/>

National Center for Biotechnology Information • National Library of Medicine • National Institutes of Health • Department of Health and Human Services

Sequence Records And The Discovery Column

Displays of sequence records in the NCBI Entrez system are enhanced to make the analysis and data access easier. The right-hand Discovery Column (A) provides direct access to analytical tools at the NCBI (B), including the "Find in this Sequence" search, and provides summaries and links to the most relevant data related to the record (C). The "Related Information" list (D) at the bottom of the Discovery Column



Display format (GenBank) and Send to menus render and save record in various formats, respectively.

provides a comprehensive set of related records from other NCBI databases.

GenBank Send to Change region shown Customize view Analyze this sequence Articles about the Pomc gene Pathways for the Pomc gene Reference sequence information More about the Pomc gene Homologs of the Pomc gene Related information

Rattus norvegicus proopiomelanocortin precursor (Pomc) mRNA, complete cds

GenBank: AF510391.1

[FASTA](#) [Graphics](#)

[Go to:](#) ☒

LOCUS AF510391 714 bp mRNA linear ROD 27-MAY-2002

DEFINITION Rattus norvegicus proopiomelanocortin precursor (Pomc) mRNA, complete cds.

ACCESSION AF510391

VERSION AF510391.1

KEYWORDS .

SOURCE Rattus norvegicus (Norway rat)

ORGANISM [Rattus norvegicus](#)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Myomorpha; Muroidea; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 714)

AUTHORS Takamori, Y. and Yamada, H.

TITLE Isolation and characterization of rat proopiomelanocortin cDNA

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 714)

AUTHORS Takamori, Y. and Yamada, H.

TITLE Direct Submission

JOURNAL Submitted (08-MAY-2002) Anatomy and Cell Science, Kansai Medical University, 10-15, Fumizono-cho, Moriguchi, Osaka 570-8506, Japan

FEATURES

source

Location/Qualifiers

1..714

/organism="Rattus norvegicus"

/mol_type="mRNA"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/tissue_type="hypothalamus"

gene

1..714

/gene="Pomc"

1..708

/gene="Pomc"

/note="prohormone; includes hormones such as MSH, ACTH, lipotropin, and endorphin"

/codon_start=1

/product="proopiomelanocortin precursor"

/protein_id="AAH43934.1"

/translation="MPRF CYSRSGALL LALLQTSIDVNSWCLSSQCDLTSTENLL ACIRACRLDLSAETVPFGNGDEQPLTENPRKYVMGHFRWDRFGPRNSSAGGSAQRRAEEETAGDGRPEPSREGKRSYSMEHFRWGVGKRRPVKVPNVAENESAFAFPL EFKRELEGEQPDGLEHVLDPDTEKADGPYRVEHFRWGNPPKDKRYGGFMTSEKSTPL VTLFKNAIIKNAHKKGQ"

ORIGIN

1 atgccgagat tctgctacag tgctcagggg gccctgctgc tggccctcct gcttcagacc

61 tccatagacg tctggagctg gtgcttggag agcagccagt gccaggacct caccacggaa

121 agcaacctgc tggcttgcat ccggccctgc agactcgacc tctcggcgga gacgcccggt

181 ttccaggcca acggagatga acagcccttg actgaaaatc cccggaagta cgtcatgggt

241 cacttccgct gggaccgctt gggcccgaga aacagcagca gtgctggcgg ctacagcgag

301 aggcgtgcgg aggaagagac ggggggggga gatggccgtc cggagccaag tccacgggag

361 ggcaagcgct cctactccat ggagcacttc cgtgggggga agccgggggg caagaagcgg

421 cgccctgtga aggtgtacc caatgtcgcc gagaacgagt gccggcaggg ctttccctta

481 gagttcaaga gggagctgga aggcagagag cctgatggct tggagcacgt cctggagcgg

541 gataccgaga aggcagagcg gccctatcgg gtggagcact tccgctgggg caaccgccc

601 aaggacaagc gctacggcgg cttcatgacc tccgagaaga gccagacgcc cctggtgacg

661 ctcttcaaga acgcatcat caagaacgag cacaagaagg gccagtggag gtgc

Change region shown **Customize view** **Analyze this sequence** **Articles about the Pomc gene** **Pathways for the Pomc gene** **Reference sequence information** **More about the Pomc gene** **Homologs of the Pomc gene** **Related information**

Change region shown and Customize view specifies regions, strand, and features displayed. Click arrow to open.

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Articles about the Pomc gene

Localization of the dopamine receptors of types 1 and 2 on [Dokl Biol Sci. 2017]

Variable proopiomelanocortin expression in tanyocytes of the [J Comp Neurol. 2017]

The expression of orexigenic and anorexigenic f [Int J Dev Neurosci. 2017]

See all...

Pathways for the Pomc gene

Endogenous sterols

Cytochrome P450 - arranged by substrate type

Phase 1 - Functionalization of compounds

See all...

Reference sequence information

RefSeq alternative splicing

See 2 reference mRNA sequence splicing variants for the Pomc gene.

More about the Pomc gene

Pomc gene

Also Known As: Pomc1, Pomc2, alphaMSH

Homologs of the Pomc gene

The Pomc gene is conserved in human, chimpanzee, Rhesus monkey, dog, mouse, chicken, zebrafish, and frog.

Related information

Protein

Taxonomy

BioSystems

Full text in PMC

Functional Class

Gene

Probe

PubMed (Weighted)

UniGene

Related citations for the Gene record from PubMed

Pathway information from NCBI Biosystems database

NCBI curated reference sequence records

NCBI Gene record

Homologous gene group from NCBI HomoloGene database

The Discovery Column provides rapid access to analysis tools and related data.

Clicking "Find-in-Sequence" opens an input box for query pattern.

Clicking "Find" updates the sequence display with identified patterns highlighted.

ATCAA

Batch Display For Sequence Records: An Example For Proteins

Multiple record views and search results in the Entrez sequence databases, such as nucleotide, protein, EST and GSS, feature rapid access to related data, flexible data display and download options, and direct access to the analysis tools (A): BLAST, the multiple alignment tool COBALT (for protein only), and the "Find in these Sequences" pattern matcher. A FASTA display of two or more protein sequence records (three in the example below) has the "Find in these sequences" function (B) activated. Searching with the input pattern, such as the one specified in ProSite syntax (C), identifies and highlights the pattern occurrences (D). You can retrieve records from other databases, related to the set of sequences displayed, by using the "Find related data" pull-down list (E).

FASTA ▾ Sort by Default order ▾

Results: 3

☐ [protein white \[Nasonia vitripennis\]](#)

1. NCBI Reference Sequence: NP_001177963.1
[GenPept](#) [Identical Proteins](#) [Graphics](#) [Related](#)

```
>gi|300388172|ref|NP_001177963.1| protein white [Nasonia vitripennis]
MATIKETPLKKFKKKVLENNQVIEAIPPINPIELQKQSDTSEKFDSDSLINKGLDQRYKQADKIAIT
YMKNNISVYCTEKNKRFYDFSTHNNINKIKQLLKNVSGIAYPGELLVIMGSSGAGKTLNINLTFRCS
DVVSGNIAINEQQVNSMMLASQMAVYQDDLFITLTVNEHLFQALVPMQKNIPYQRIIRRVNEVISE
LALTLCRNTIIGIPKIKGISGGEMKRLSFASEVLTDPPLMLRDEPTSDLSYMAHQVSVLKNLTGRGK
TIITTLHQPSSELSFMSXILLTAEGRVAFMGSTEEAEFFKSLGASCPNNVNPADFYQLLAVVPGREY
ACRHAHKKVCDQFECNSLGGKISNKKVQNIIDFVEENHYRNKMMQIKNTYKASWLEQKAVLWRSWLSVI
KEPILIKVRLQTIHVSLLIGIYLNQKLDQDGMNININGALFIFLTMTFQNVFAVHVFCSELPFLRE
HANGMYRTZYFICKTAAEAPFIAVPLIFTIAYPMIGLYPDIRHFFTTVGIVLVANVSTSGYLISC
TSTHLSMVLISGPPVPIFLFGGFFLNTASVPIYKWFSLHFRYGNELLNQADIESINCTRSNV
TCPKSGMLVQLTLNFKOFFWMDICALVVLIIIVFRVSVLLALISCTFKKRN
```

☐ [cystic fibrosis transmembrane conductance regulator \[Rattus norvegicus\]](#)

2. NCBI Reference Sequence: NP_113694.1
[GenPept](#) [Identical Proteins](#) [Graphics](#) [Related Sequences](#) [Identical Proteins](#)

```
>gi|91982740|ref|NP_113694.1| cystic fibrosis transmembrane conductance regulator [Rattus norvegicus]
MQKSPLEKASFISKLFSSWTPILRKGYRHHLELSDIYQAPSSDSADHLEKLERENDREQAKKKKPOLI
HALRRCFVNRFFVGYLLYLGEVTKAVQPVLLGRITASYDPDNTTEERSIATYLGIGLCLLFTVRTLHP
ATFGLHHIGQMRIAMFSLIYKTKLSSRVLDKISGOLISLLSNLNKFDGLALAHFIHIAPILOVVL
LMGLLMDLLQSFACGLGLLIVLVIFQAILGKMMVYRDRAAKINERLVTSEVIDNIYSVKAYCWESA
MEKIESLREEELKHTRRASAYMRFTSSAFFSGGFVFLSVLPYTVINGIVLRKIFTTISFCVLRMSV
TRQFPTAVQIWDYSLGIRKIQDFLQTEYKVLVNLMTGLVMENVTAFAWEGFQELLEKVQLNDDRK
TSNGENHLSPFSLCLVGNPVLKNNLNKKGEMAITGSGAGKTSLLMLTLGELEASEGIIKSHGRVSF
SQISWMIPGTIKENIIFGVSDEYRYKSVKAKQLQEDITKFAEQDNTVLGEGGVTLGGQRRARISLAR
AVYKDADLYLLDSPFGYLDVLTTEEQIFESCVCMLMASKTRILVTSKMEQLKADKILILHEGSSYFYGT
SELQSLRDPFSSKLNGYDTFDQFTEERRSSITETLRRFSVDDASTWNKAKQSFRTGEFGEKRNKNSIL
SFSFSSVKKISIVQKTPLSIEGESDQLQERRISLPDSEHGAEALPRSNMITAGPTFFGRRRQSLDLMFT
TPSSVSSSLQRTASIRKISLAPRISLKEGDIYSRRLSQDSTLNITEEINEEDLKECFDDMMVKIPTVT
WNTYLRVFTLHRLGFAVLINCVLVFLVEVAASLFWLLKNNPVGNGNGTKIANTSYVYVITSSSFYI
FYIYGVADTLALSLFRGLPVHTLITASKILHRKMLHSILHAPMSTFNKLAGGILNRFSDIAILD
FLPLTFIDFQLLFVVGAIIVVSALQFYIFLATVPLGAVFLLRAYFLHTSQQLQLESEGRSPFIFTH
VTSLLGLNLTARFRQTYFETLFHKAALNHTANMFMYLATLRFQMRIDMTFVLFFIVVTFISLITGEG
EGTGTIILTLAMNISTLQMAVNSSIDTSLMRSVSRVFKFIDIQTESEICTKIMKELHSED
NEHVKKCDTPSGGEMVYKDLTVKYDDGNAILNISFSISPGORVLLGRTGSGKSLTSLA
GETIQDGVSNMNTLQENKAFGVYTKVFIFSGTERQHLDPNGKWRDEEINKVADQVGLKS
LNFTLVGGYVLSHGKQLMCLARSVLSKAKIILLDEPSANLDPITYQVIRRVLRQAFAGCT
EAMLDQCQRLVIEQGNVMQYESLQALLSEKSVFQALSSSEKMKLFHGRHSSQKQKPTQITA
VQETRL
```

☐ [DNA mismatch repair protein Mlh1 isoform 1 \[Homo sapiens\]](#)

3. NCBI Reference Sequence: NP_000240.1
[GenPept](#) [Identical Proteins](#) [Graphics](#) [Related Sequences](#) [Identical P](#)

```
>gi|4557757|ref|NP_000240.1| DNA mismatch repair protein Mlh1 isoform 1 [Homo sapiens]
MSFVAGVIRRLDETVMVRIAGAEVIRPANAIEKEMENCLDAKSTSIQVIKVEGGLKLIQIQDMGTGIRK
EDLDVCEFTTSKLSFDEIASISTYGRGEALASISHVAHVITTTKTAGDKCAYRASYSOGKLUKAPPK
PCAGNQGTQITVEDLFYNIATRRKALKNPSEEGKILEVVGGRYSVHAGISFSVKKQGETVADVRLPNA
STVDNIRSFIGNAVRELIEIGCEDKTLAFKMNIGYTSANYSVKKCIFLLFINHRLVESTSLRKAIEIVY
AAYLPKNTHPFLVLSLEISQNVQVDMVHPTKHEVHFLHEESILERVQKHIESKLLGSSNRMVYFTQTLPL
GLAGPSGEMVKSSTSLTSSSTSGSSDKVYAHQMVRTDSREQLDLAFLQPLSKPLSSQQAIVTEDKTDIS
SGRAAQDDEFPPAPAEVAAKNQSLGDDTTKGTSEHSEKRGPTSSNPKRHREDSVEMVEDDSRKEH
TAACTPRRRVSVLSLQEEINEQGHEVLRMLHNSFVGVNPNQWALAQHQTLYLLNTTKLSEELF
YQILTYDFANVLRSLSEAPLFDLAMLALDPSGWTEDGPKGLEAEYIEVFLKKKAEMADYFSLIEI
DEEGNLTLGLPLIDNYVPPIELPLIFILRLATEVWDEEKECFESLSKECAMFYSIRKQYISEESTLSGQ
QSEVPGPSIPSNKWTVEHIVYKALRSHILPPKHFTEDGNILQLANLPDLKVKVFERC
```

Send to: ▾

Filters: [manage filters](#)

Results by taxon

Top Organisms [\[Tree\]](#)

- Rattus norvegicus (1)
- Homo sapiens (1)
- Nasonia vitripennis (1)

Analyze these sequences (A)

- Run BLAST
- Align sequences with COBALT
- Identify Conserved Domains with CD-Search
- Find in these sequences (B)

Find related data (E)

Database: Select

- Select
- Biocollections
- BioProject
- BioSystems
- Conserved Domains
- Gene
- Genome
- GSS
- HomoloGene
- Nucleotide
- OMIM
- PMC
- PopSet
- Protein
- Protein Clusters
- PubChem BioAssay
- PubChem Compound
- PubChem Substance
- PubMed
- SNP

Find related data

Database: Nucleotide

Option: Encoding mRNA

Link from protein to encoding mRNA

Find items

Autonomous protein records

- Nucleotide
- Mature Peptide
- Order cDNA Clone
- Encoding mRNA
- All nucleotide sequences from this genome
- TSA project
- WGS Project
- Genomic records

Find (C)

1 of 4

NP_001177963 : 121-128

Help

Feature Highlight: Correlating Subsequences To Their Annotation Feature

The GenBank flatfile display of sequence records shows the locations of biological features in the FEATURES section (A). A new "Feature highlight" function was added to relate functional annotations with the sequence. In the GenBank report with the FEATURES section displayed, clicking any linked feature (B) highlights the corresponding region in the sequence (C) and opens the feature highlight bar (D) at the bottom of the record.

GenBank Send to: ▼

Homo sapiens complement factor H (CFH), RefSeqGene (LRG_47) on chromosome 1

NCBI Reference Sequence: NG_007259.1
[FASTA](#) [Graphics](#)

[Go to: ▼](#)

LOCUS NG_007259 102494 bp DNA linear PRI 26-NOV-2018
 DEFINITION Homo sapiens complement factor H (CFH), RefSeqGene (LRG_47) on chromosome 1.
 ACCESSION NG_007259
 VERSION NG_007259.1
 KEYWORDS RefSeq; RefSeqGene.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Mammalia; Eutheria; Euarchontariida; Catarrhini; Hominidae; Homo.
 COMMENT REVIEWED REFSEQ: This record has been curated by NCBI staff in collaboration with the RefSeq project. This sequence was derived from a combination of genomic DNA and complementary DNA. This sequence is a reference sequence.

Summary: This gene is a member of the Regulator of Complement Activation (RCA) gene cluster and encodes a protein with twenty short consensus repeat (SCR) domains. This protein is secreted into the fluid phase and is involved in the regulation of the complement system.

FEATURES

Feature	Location/Qualifiers
source	1..102494 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="1" /map="1q31.3"
gene	4868..100494 /gene="CFH"
exon	4868..5165 /gene="CFH" /gene_synonym="AHUS1; AMBP1; ARMD4; ARMS1; CFHL3; FH; FHL1; HF; HF1; HF2; HUS" /inference="alignment:Splign:1.39.8" /number=1
variation	4872 /gene="CFH" /gene_synonym="AHUS1; AMBP1; ARMD4; ARMS1; CFHL3; FH; FHL1; HF; HF1; HF2; HUS" /replace="c" /replace="t" /db_xref="dbSNP:898576996"
CDS	join(5108..5165,25968..26153,26847..26952,28979..29055,30466..30657,32613..32783,38054..38227,42410..42604,43053..43229,66725..66907,68583..68759,78111..78287,79460..79642,79751..79930,81336..81512,89814..89996,90465..90650,93609..93782,94865..95041,96442..96618,98807..98989,100101..100303) /gene="CFH"

Feature Highlight Bar (D):

Feature: **CFH** (4868..5165)
 /gene="CFH"
 /gene_synonym="AHUS1; AMBP1; ARMD4; ARMS1; CFHL3; FH; FHL1; HF; HF1; HF2; HUS"
 /inference="alignment:Splign:1.39.8"
 /number=1

FASTA:

```

4861 gccggtcaca gcattaacat ttagtgggag tgcagtgaga attgggttta acttctggca
4921 ttcttgggct tgggcttgt gttgatttt ttatttactt tgcaaaagt tctgatagcc
4981 ggagcatcta gtttcaactt ctttttcag caagtcttt cctgcactaa tcacaattct
5041 tgggaaggga gaactggagc ttgtgaacag agttagctgg taaatgtcct cttaaaagat
5101 ccaaaaaatg agacttctag caaagattat ttgccttatg ttatgggcta ttgtgtatgc
5161 agaaggttag attaaaaagag attttctt gaaactgta ttatgaaaca ttgtctaagt
5221 atgcttttca caggagtaat atttttga tttagaaaat gtgcttaagt attctgtaac
  
```

Annotations:

- A:** FEATURES section
- B:** Find in this Sequence link
- C:** Highlighted region in the sequence
- D:** Feature highlight bar
- E:** Feature pull-down list
- F:** Feature link in the menu bar
- G:** Navigation arrows
- H:** Details box
- I:** Deactivate function icon

You can switch the features highlighted to other types using the pull-down list (E). The "Feature" (F) link in the menu bar anchors the display to the corresponding item in the FEATURES section. Arrows (G) allows the navigation through the features in the selected category. The number of highlighted feature and the total number are shown between the two set of arrows. The Details box shows the annotation for the highlighted feature, which you can minimize using the down arrow (H). FASTA and GenBank links display the highlighted subsequence containing the feature in these formats. Deactivate the function and removes the controls by clicking the "x" icon (I).

Other Display Format And Download Options

You can display and download sequence records in several formats, such as Summary, GenBank, *etc.*, using pull-down menus linked to display formats (A) and “Send to” (B). The “Revision History” (C) option tracks changes of the sequence record. This specialized display allows direct pairwise-comparison of different versions of a record to highlight sections where changes or updates occurred.

Summary

Send to:

Homo sapiens complement factor H (CFH), RefSeqGene (LRG_47) on chromosome 1

102,494 bp linear DNA

Accession: NG_007259.1 | GI: 16396538

Protein PubMed Taxonomy

GenBank FASTA Graphics

Format

- ☒ Summary
- ☐ GenBank
- ☐ GenBank (full)
- ☐ FASTA
- ☐ FASTA (text)
- ☐ Graphics
- ☐ ASN.1
- ☐ Revision History
- ☐ Accession List
- ☐ GI List

GenBank

Homo sapiens complement factor H (CFH), RefSeqGene (LRG_47) on chromosome 1

NCBI Reference Sequence: NG_007259.1

FASTA Graphics

Complete Record

Coding Sequences

Gene Features

Choose Destination

- ☒ File
- ☐ Clipboard
- ☐ Collections
- ☐ Analysis Tool

Download 1 item.

Format

FASTA

Show GI ☐

Create File

Choose Sequence Analysis Tool

- ☒ BLAST
- ☐ PrimerBLAST

Submit

Additional formats, such as FASTA, Feature Table (for features annotated on the record), and XML, are available through the “Send to” menu (D). For nucleotide sequences with annotated coding regions, the FASTA sequences of these regions can also be downloaded using the “Coding Sequences” (E) option. Selecting the “Analysis Tool” (F) option expands the “Send to” menu to include options for analysis tools (G). From here, you can submit the sequences to BLAST search or to target-specific primer design using Primer-BLAST.

Example: Coding Sequence Download

A common task is to get the coding sequences for all of the genes annotated on a bacterial genome or plasmid. The displays below show the partial results of downloading FASTA nucleotide coding sequences (H) and their encoded protein sequences (I) for the Escherichia coli plasmid p271A (www.ncbi.nlm.nih.gov/nuccore/NC_015872.1) using the “Send to >> Coding Sequences” option (J).

Complete Record

Coding Sequences

Gene Features

Download features.

Format

FASTA Nucleotide

Create File

FASTA Nucleotide

FASTA Protein

```
>lcl|NC_015872.1_cds_YP_004765012.1_1 [gene=nuc] [protein=plasmid conjugative transfer
endonuclease] [protein_id=YP_004765012.1] [location=complement(870..1235)]
ATGATGGCTTATAGCTTTACCGCGATGGATATAATGCAGGCGCTGGTCAAAGCTAAAAATCGCGGTGTCG
ATGTGAAAAATCGTTATCGACGAACGCGGCAATCGTAATGACGTAAGCCCGTGGATGAAGTACATTGT
GAAAAACGAGATCCGCTGCGGGTGGATAGCGACTTCTCTATCCAGCATGACAAGGTGATGATTATGAT
GAACGGTCAAGTCAGACGGGCGAGTTCACCTACACCAAGAGCAGCAGAA
TGGTGATCTGGAACATGCCAGCTGGCAAAACCTTATCTTGAGCACTC
GGAGGACTATAAATGA
>lcl|NC_015872.1_prot_YP_004765012.1_1 [gene=nuc] [protein=plasmid conjugative
transfer endonuclease] [protein_id=YP_004765012.1] [location=complement(870..1235)]
MMAYSFTAMDIMQALVKAKNRGVVDKIVIDERGNRNDVSHRAMKYIVKNEIPLRVSDSDFPIQHKVMKIID
ERSVQTGSFNYTKAAESKNSENAVVIWNPQLAKPYLEHWEDRWNRGEDIYK
>lcl|NC_015872.1_cds_YP_004765013.1_2 [gene=traG] [protein=TraG]
ATGATCAATCATCGGCTTATTACCAAGTGGGCGCGTGGCGGAGTATT
AAGTCCGAATCAACCGTTATAACGAGGTGGTTTGTGACACATTGCAAGC
CAGCATTCACCGCGCGTTTATCGATAACCTGACTGCGGCATTAAACAGT
ATGATTAATGACGTGGTGTTCACGGAGCGTTCAGGGGAATTATCTGTT
ACACCGCGCTATCGCTTTTCGTAATAAATATCGCGCTGGATAAAGACCT
TATCTTGTAGTTTTGCCGGGAATTGACAGCCATAAGCGCGAGCTGGAC
GAGTTATACCGGGCTAAACCGCGGGAAGAAATTTTGTCTACTGCGGTAT
TTGTGGGAGAAACCGGCTCAGGGAACCGTACTACCCGCGCGCTGCT
ACGCGTGATAATCATGGAGGATGCTCAGCAAGTTGAAGCTACCCACCTC
TATGGCGGTAAGATAAACCAGGGCTGGCTACGCGCAGCAGTGCCTT
CAGATCGTATCTTCATGACAGAGCTGAGGGATGATGCTGCATGGGATT
ACACCGTGGCGGCTTAACCTCCACACATGCTAACTCGGCGATGGATGCT
GTTAAAGCTACCGAAGTGGGCGTATGCTGGATATTTCTGACATTATG
ATGTGCTGCTTTTTCATGGCGAAAAGGAGATTGCGCAAACTACTATTG
CATGAATGGAGGATAAAGTGA
```